Stochastic Processes and their Applications 34 (1990) 155-169
North-Holland

A DYNAMIC STORAGE PROCESS

C. KIPNIS

Centre de mathematiques appliquees, Ecole polytechnique, 91128 Palaiseau, France

Ph. ROBERT

Institut National de Recherche en Informatique et en Automatique, Domaine de Voluceau, 78153 Rocquencourt Le Chesnay, France

Received 12 August 1987
Revised 22 June 1988

We analyse a storage process with dynamical arrivals and departures. Under probabilistic assumptions, we study the behavior of the storage unit and give the main features of it: the maximal throughput and the occupied length of the unit.

dynamic storage allocation * maximal throughput * measure valued Markov processes

Introduction

The phenomena of fragmentation which occurs in many storage models is a well known problem. For example if the storage unit is a disk cylinder and the items to be packed are records or files (we assume that they are stored only at one address), after some arrivals and departures of the items in the unit, the free space of the unit is then made of small portions which are useless although the concatenation of these spaces could be usable. Thus one of the important questions in this area is to know if and when one should compact the storage area given that this operation has a certain cost.

The results concerning dynamic storage are quite rare. The main problems in this area are presented and discussed in Benes (1982) (see also Aven, Coffman and Kogan, 1987). In the case of an infinite storage unit, fragmentation has been analysed in Coffman, Kadota and Shepp (1985). The authors assume that the items take the first free space on the unit and that the unit is not compacted (see also Aldous, 1986). In Coffman, Garey and Johnson (1983) a comparison of algorithms in the worst case is done. In this case there are many storage units and the goal is to use a minimum number of units given arrivals and departures of items. The problem of periodically reorganizing the storage unit has been analysed in Scholl (1979). The compression of the unit in this case is done when the address of an item exceeds a given value.

0304-4149/90/$3.50

Our purpose is to study what can be gained with the compression of the storage unit. We assume that the size and the residence time in the unit of the items have a statistical distribution and as soon as there is a hole in the unit it is immediately compacted. In order to evaluate this compression policy we analyse the main quantities of interest: the maximal throughput and the occupation of the storage unit. Our main tool is the description of our storage process as a measure valued Markov process which seems quite relevant for this analysis. Most of the models previously analysed are reduced to a one or two dimensional process. Our approach is also an attempt to give a natural (although not easy to handle) framework to these problems in order to tackle them. Finally we must add that we have used the symbolic computation package MACSYMA to guess the explicit formula of Theorem 4.1 and that Section 7 owes much to J.L. Bouchenez and M. Loyer's SM90 "speedy".

1. The model

In the present paper we start a study of a dynamic storage process, which can be described as follows: we have a storage unit (the bin) of finite capacity (say one); items arrive at rate α to be stored which require random independent identially distributed room. Each item is stored immediately if there is room enough, otherwise it remains waiting until it can be accommodated. Items are stored in their order of arrival so that one item waiting forces all the subsequent arriving items to wait. Each item after having entered the storage unit remains there for an exponential holding time of parameter $p(x)$ depending only on its size x. At one departure the storage unit is recompacted so that no room is lost and then the first waiting item is immediately stored provided the empty space in the bin exceeds its size, otherwise it waits for the next departure.

In this context several quantities are of interest: The maximum throughput of the system which is the biggest arrival rate that does not generate in the long run an infinite waiting line before the bin, the number of items that are stored and the proportion of the bin that is used under this optimal policy. Clearly enough these quantities can be computed in terms of the equilibrium distribution of the state of the bin for the auxiliary system obtained by supposing that initially an infinite number of items is already waiting for service provided one has an ergodic theorem.

The paper is organized as follows: In Section 2 we set some notations and study the construction of the process as well as some intuitive results on the output of the process. In Section 3 we prove existence and uniqueness of the invariant measure associated with this process. In Section 4 an explicit formula for the maximal throughput is given in terms of the behavior of the random walk generated by the sizes of the items. From this we derive in Section 5 several bounds based on simple functionals of the sizes of the items. In Section 6 we present some examples and also some additional results in simple cases. Finally Section 7 is devoted to the numerical applications of our results.

2. Notations and construction of the process

In view of the introduction we will describe the set of items inside the bin by a (random) point measure on $[0, 1]$ i.e. an element $\eta \in M$ of the form $\eta(\mathrm{d}x) = \sum_y \delta_y(\mathrm{d}x)$ (δ_a is the Dirac mass in a), with a positive but finite number of points in the sum. We will write $x \in \eta$ if x satisfies $\eta(\{x\}) > 0$. This space is naturally endowed with the weak topology on measures i.e. the topology that makes the application $\eta \to \langle f, \eta \rangle$ continuous for all continuous functions f on $[0, 1]$, with the shorthand notation

$$\langle f, \eta \rangle = \int f(x)\eta(\mathrm{d}x).$$

For example $\langle 1_{[a,b]}, \eta \rangle$ ($1_{[a,b]}$ is the indicator function of $[a, b]$), gives the number of items in the bin with size between a and b. All the relevant quantities can be expressed as such integrals. In particular the occupied length of the bin is given by $\langle x, \eta \rangle$, the number of items by $N = \langle 1, \eta \rangle$ and the processing rate in the state η by $\langle p, \eta \rangle$.

We will use the notation P for the probability measure on $U = ([0, 1] \times \mathbb{R}_+)^N$ which is the product measure of $p(x)\,\mathrm{e}^{-p(x)u}\mu(\mathrm{d}x)\,\mathrm{d}u$ and $(X_i, T_i)_i$ will denote the coordinate random variables, respectively the size of the items and their residence time. Thus X_1 has the distribution μ and T_1 is exponentially distributed with parameter $p(x)$. The shift operator S on U is defined by $S((x_i, t_i)_i) = (x_{i+1}, t_{i+1})_i$.

Our dynamic storage process is completely described by $(\eta_t, R_t)_{t \geq 0}$, with $\eta_t \in M$ being the state inside the bin and R_t the size of the first waiting item at time t. Its generator Ω is defined by the way it acts on bounded continuous functions F on $M \times \mathbb{R}_+$:

$$\begin{aligned} \Omega(F)(\eta, r) = \sum_{x \in \eta} p(x) \times \Bigg[E\Bigg(\sum_{k=0}^{+\infty} F\Bigg(\eta - \delta_x + \delta_r + \sum_{i=1}^{k} \delta_{X_i}, X_{k+1} \Bigg) \\ \times 1_{\{r + \langle 1, \eta \rangle - x + \sum_{i=1}^{k} X_i \leq 1 < r + \langle 1, \eta \rangle - x + \sum_{i=1}^{k+1} X_i\}} \Bigg) \\ + F(\eta - \delta_x, r) 1_{\{\langle 1, \eta \rangle + r - x > 1\}} \Bigg] - \langle p, \eta \rangle F(\eta, r), \end{aligned}$$

where 1 in $\langle 1, \eta \rangle$ denotes the constant function 1 and $\sum_{i=1}^{0}$ is by convention equal 0.

Proposition 2.1. *The process described by the generator Ω is uniquely defined.*

Proof. Since we are dealing with a jump process, the only obstruction to uniqueness is explosion i.e. that in a finite time an infinite number of items enter the bin. To see that this is impossible, we define by induction the non-decreasing sequence $(U_n)_{n \geq 0}$ such that

$$U_0 = 0, \qquad U_n = \inf\left\{ t \,\middle|\, X_n + \sum_{i=0}^{n-1} X_i 1_{\{U_i + T_i > t\}} \leq 1 \right\}. \tag{2.1}$$

Note that U_n is the time at which the nth item is packed in the bin. Our storage process is then defined by

$$\eta_t = \sum_{i=0}^{\infty} 1_{\{U_i+T_i>t\geq U_i\}}\delta_{X_i}, \qquad R_t = X_n \quad \text{if } U_{n-1} \leq t < U_n.$$

Now it is sufficient to prove that $\lim_{n\to+\infty} U_n = +\infty$. (Here and in the rest of the paper all the relations involving random variables are to be understood almost surely.) Recall that we assume $\mu\{0\} = 0$ so there exists $p > 1$ and a subsequence $(X_{n_k})_k$ such that $X_{n_k} > 1/p$. Now we remark that

$$U_{n_{kp}} \geq U_{n_{(k-1)p}} + \inf\{T_{n_{(k-1)p+i}} \mid 0 \leq i \leq p-1\},$$

because the item of index n_{kp}can be packed in the bin only if one of the $(X_{n_{(k-1)p+i}})_{0\leq i<p}$ has left the bin.

Using that $(\inf\{T_{n_{(k-1)p+i}} \mid 0\leq i\leq p-1\})_{k\geq 1}$ are i.i.d. and the law of large numbers, we get that $U_{n_k} \to +\infty$ P-a.s. and our proposition is proved. □

We now prove an intuitive and useful coupling lemma.

Coupling Lemma 2.2. *Suppose that* (X_n, T_n) *and* (Y_n, K_n) *satisfy almost surely for all n,* $X_n \geq Y_n$ *and* $K_n \leq T_n$ *then also* $U_n \geq V_n$ *where* U_n *(resp.* V_n*) are defined by* (2.1) *for the sequences* (X_n, T_n) *(resp.* (Y_n, K_n)*).*

Proof. Let us denote by $u_1, \ldots, u_n, \ldots,$ the entrance times in the system of item number n in the (X, T) process and $v_1, \ldots, v_n, \ldots,$ those corresponding in the (Y, K) process. Then clearly it is enough to prove that almost surely $u_n \geq v_n$ for all n.

The claim is true for $n = 0$; Denote by N the first index n such that $u_n < v_n$. Suppose that N is finite then v_N, the entrance time of X_N, is equal to one of the $v_k + K_k$ with $0 \leq k < N$. But since $K_k \leq T_k$ we have for all $k < N$, $v_k + K_k < u_k + T_k$ by induction hypothesis and the inequality between the X's and Y's. Therefore all the items that have left the processor for the X-process have already left also for the Y-process and the items that remain are smaller. Therefore there is room enough to accomodate the Nth item also in the Y-process. This is a contradiction and, our lemma is proved. □

Proposition 2.3. *The sequence* $(n/U_n)_{n\geq 0}$ *converges a.s. to a constant* $\rho(\mu, p)$ *called the maximal throughput of the storage process.*

Proof. If $\omega = (x_n, t_n)_{n\geq 0}$ then

$$U_{n+p}(\omega) = U_n(\omega) + U_{k_1+p}((y_1, s_1), \ldots, (y_{k_1}, t_{k_1}), (x_n, t_n), (x_{n+1}, t_{n+1}), \ldots)$$

where $y_1, \dots, y_{k_1}$ (resp. $s_1, \dots, s_{k_1}$) are the sizes (resp. residual processing times) of the remaining items in the bin when X_n is packed. But

$$\begin{aligned} &U_{k_1+p}((y_1, s_1), \dots, (y_{k_1}, s_{k_1}), (x_n, t_n), (x_{n+1}, t_{n+1}), \dots) \\ &\quad \geq U_{k_1+p}(0, \dots, 0, (x_n, t_n), (x_{n+1}, t_{n+1}), \dots) \end{aligned}$$

according to the coupling lemma applied to the two sequences $((y_1, s_1), \dots, (y_{k_1}, s_{k_1}), (x_n, t_n), (x_{n+1}, t_{n+1}), \dots)$, $((0,0), \dots, (0,0), (x_n, t_n), (x_{n+1}, t_{n+1}), \dots)$, the right term of this inequality is simply $U_p((x_n, t_n), (x_{n+1}, t_{n+1}), \dots)$ or $U_p(S^n(\omega))$ with S^n the nth iterate of the shift S.

Finally we get that

$$U_{n+p}(\omega) \geq U_n(\omega) + U_p(S^n(\omega)).$$

The sequence $(U_n)_{n\geq 0}$ is thus superadditive and the shift S is associated with a product measure so Kingman's superadditive theorem (cf. Neveu, 1983) ensures the a.s. existence of a limit for the sequence $(U_n/n)_{n\geq 0}$. □

We now prove an intuitive result on this throughput.

Proposition 2.4. *If μ is stochastically larger than v and* $\inf\{q(y) \mid y \leq x\} \leq p(x)$ *then* $\rho(\mu, p) \leq \rho(\nu, q)$.

Proof. Since $\mu \geq^{\mathrm{st}} \nu$ we can couple the two probabilities on $[0, 1]$, that is we can construct a measure Q on $[0, 1]^2$ such that its first (resp. second) marginal is μ (resp. ν) and that it is concentrated on $\{(x, y) \mid x \geq y\}$.

Denote now by $(X_1, Y_1), \dots, (X_n, Y_n), \dots$, a sequence of i.i.d. random variables with distribution Q. In particular it satisfies $X_n \geq Y_n$ a.s. for all $n \geq 0$.

Besides we can choose a sequence of i.i.d. exponential (mean one) random variables $(\tau_n)_{n\geq 0}$ and set

$$T_n = \frac{1}{p(X_n)}\tau_n \quad \left(\text{resp. } S_n = \frac{1}{q(Y_n)}\tau_n\right)$$

for the residence time of the X_n (resp. Y_n). The result follows from the coupling lemma. □

3. Existence and uniqueness of a stationary probability

In order to prove existence and uniqueness of an invariant measure for our process we will rely on the theory of ϕ-irreducible Markov chains (see Revuz, 1975).

A natural measure that will appear in the sequel is $\phi(\mathrm{d}\eta, \mathrm{d}r)$ defined as follows; for F a bounded measurable function on $M \times \mathbb{R}_+$,

$$\int F(\eta, r)\phi(\mathrm{d}\eta, \mathrm{d}r) := \sum_{n\geq 1} E\left(F\left(\sum_{i=0}^{n} \delta_{X_i}, X_{n+1}\right) 1_{\{S_n \leq 1 < S_{n+1}\}}\right)$$

with $S_n = \sum_{i=1}^{n} X_i$, which is the measure giving the distribution of (η, R) starting with an empty bin, immediately after time 0.

We will work with the discrete time Markov chain (the skeleton of our continuous time Markov process) obtained by considering the state of the process at each departure time. Its successive states will be denoted by $(\eta_n, R_n)_n$. Its transition probability $P((\eta, R), (\mathrm{d}\eta, \mathrm{d}r))$ satisfies the following easy estimate.

Set $\Delta_q^n = \{\gamma \in M \mid \langle 1, \gamma\rangle \leq q \text{ and } \langle p, \gamma\rangle \leq n\}$ and $\Delta_q = \Delta_q^{\infty}$. If $\langle 1, \eta\rangle = k$ then there exists a constant $\alpha(k, q, n, \eta, R)$ such that

$$P^{k+1}((\eta, R), (\mathrm{d}\gamma, \mathrm{d}r)) \geq \alpha\phi(\mathrm{d}\gamma, \mathrm{d}r) 1_{\Delta_q^n}(\gamma, r) \quad \text{on } \Delta_q^n. \tag{3.1}$$

This result follows from observing that the probability that all the $k+1$ elements of (η, R) are processed before any new item might be processed is bounded away from 0. In particular when the function p is bounded by a constant c, the bound α is uniform for $(\eta, R) \in \Delta_k$ and is larger than

$$\int_{\mathbb{R}^2} c \frac{(cx)^k}{k!} \mathrm{e}^{-cx} 1_{\{y-x>0\}} n \, \mathrm{e}^{-ny} \, \mathrm{d}x \, \mathrm{d}y.$$

Inequality (3.1) ensures ϕ-irreducibility which we recall means that for any set A such that $\phi(A) > 0$, for any (η, R) there exists a n such that $P^n((\eta, R), A) > 0$. We will now prove one result on existence and uniqueness under the hypothesis that the items leave the bin with a minimum rate.

Theorem 3.1. *If the function $p(x)$ is bounded below by $c > 0$ then there exists a unique invariant probability measure $m(\mathrm{d}\eta, \mathrm{d}r)$ for the continuous time process. Moreover the process is ergodic and for any bounded measurable function F, we have P-a.s.,*

$$\lim_{T \to +\infty} \frac{1}{T} \int_0^T F(\eta_s, R_s) \, \mathrm{d}s = \int F(\eta, r) m(\mathrm{d}\eta, \mathrm{d}r).$$

Proof. Because of the ϕ-irreducibility, the Markov chain is either transient or recurrent.

If it were transient there would exist an increasing sequence of measurable subsets F_n such that $M \times \mathbb{R}_+ = \bigcup_{n=1}^{+\infty} F_n$ and for all (η, R) we would have

$$G(\eta, R; F_k) = \sum_{n=1}^{+\infty} P^n((\eta, R), F_k) < +\infty.$$

Now given any $\Delta_{q_0}^n$, there exists a k such that $F_k \cap \Delta_{q_0}^n = A$ satisfies $G(\eta, R; A) < +\infty$ and $\phi(A) > 0$.

But this implies that for all p we have

$$+\infty > G(\eta, R; A) \geq G(\eta, R; \Delta_q) \alpha_{q,q_0} \phi(A)$$

since in the particular case where p is bounded below the bound α in (3.1) becomes uniform on Δ_q as noticed previously. Hence any Δ_q is visited only a finite number of times. Therefore, since on Δ_q the total rate $\langle p, \eta\rangle$ is larger than $q \times c$, we must have that almost surely $\lim_{n\to+\infty}\langle p, \eta_n\rangle = +\infty$. But this clearly implies that for our continuous time process the mean throughput is a.s. infinite which is a contradiction with Proposition 2.3. Therefore our Markov chain is recurrent and there exists a unique invariant measure m (up to a multiplicative constant) for the discrete time skeleton.

We now want to exclude the case of null recurrence that is, when m has an infinite mass. If this were the case, one would have as above a set A such that $A \subset \Delta_{q_0}$ and $0 < m(A) < +\infty$. By irreducibility we also have $\phi(A) > 0$. Hence for all k,

$$+\infty > m(A) = mP^{q_0+1}(A) \geqslant m(\Delta_k)\alpha_{k,q_0}\phi(A) > 0.$$

Therefore by Jain and Jamison's theorem (Revuz, 1975), $P^n((\eta, R); \Delta_k)$ tends to zero as $n \to +\infty$. But in this case

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n=1}^{N} P^n\left((\eta, R); \bigcup_{i=1}^{k} \Delta_i\right) = 0$$

so that for any k the proportion of time spent by η_n in $\bigcup_{i=1}^{k} \Delta_i$ tends to 0. This also implies that the throughput is infinite.

Thus the discrete time skeleton has a unique invariant probability measure $\nu(\mathrm{d}\eta, \mathrm{d}r)$ and the (unique) invariant measure for our process is then proportional to

$$\frac{1}{\langle p, \eta\rangle}\, \nu(\mathrm{d}\eta, \mathrm{d}r).$$

This measure has a finite mass since $\langle p, \eta\rangle \geqslant c$ if $\eta \in M$. The last assertion of the theorem is an application of the ergodic theorem for recurrent Markov chains (cf. Revuz, 1975). □

Corollary 3.2. *Under the assumptions of Theorem* 3.1, *for any bounded Borel function* f *on* $[0, 1]$,

$$\int \langle f, \eta\rangle m(\mathrm{d}\eta) = \rho(\mu, p) \int_0^1 \frac{f(x)}{p(x)}\, \mu(\mathrm{d}x).$$

Proof. For any $t > 0$,

$$\frac{1}{t}\int_0^t \langle f, \eta_s\rangle\, \mathrm{d}s = \frac{1}{t}\sum_n f(X_n) T_n 1_{\{U_n+T_n\leqslant t\}} + \frac{1}{t}\sum_n f(X_n)(t - T_n) 1_{\{U_n\leqslant t< U_n+T_n\}}.$$

It is easily seen that the last term converges a.s. to 0 as t goes to $+\infty$, the equality is then a consequence of Theorem 3.1 for the left-hand side and of the law of large numbers and the definition of ρ_μ for the right-hand side. □

In particular the mean occupied length of the unit in steady state is

$$\rho_\mu \int_0^1 \frac{x}{(p(x)}\, \mu(\mathrm{d}x).$$

4. An explicit formula for the constant rate

For the remaining paragraphs we will assume that the processing rate $p(x)$ is constant (say one) and use $\rho(\mu)$ instead of $\rho(\mu, p)$. We have the following result.

Theorem 4.1. *The invariant probability measure m for the process* (η_t, R_t) *is given by the formula, for any bounded measurable F function on* $M\times\mathbb{R}_+$,

$$\int F(\eta, y)m(\mathrm{d}\eta, \mathrm{d}y) = c(\mu)\sum_{n=1}^{+\infty}\frac{1}{n}E\left[F\left(\sum_{i=1}^{n}\delta_{X_i}, X_{n+1}\right)1_{\{\sum_{i=1}^{n}X_i\leqslant 1<\sum_{i=1}^{n+1}X_i\}}\right]$$

where $c(\mu)$ *is a normalizing constant.*

Proof. Recalling the form of the generator, it is enough to prove that $m\cdot\Omega=0$ i.e. for any function F, to prove the identity

$$\sum_{n=1}^{+\infty}\frac{1}{n}\sum_{i=1}^{n}\sum_{p=0}^{+\infty}E\left[F\left(\sum_{j\neq i,1\leqslant j\leqslant n}\delta_{X_j}+\sum_{j=n+1}^{n+p}\delta_{X_j}, X_{n+p+1}\right)\right.$$
$$\left.\times 1_{\{\sum_{j=1}^{n}X_j\leqslant 1<\sum_{j=1}^{n+1}X_j,\sum_{j\neq i,1\leqslant j\leqslant n}X_j+\sum_{j=n+1}^{n+p}X_j\leqslant 1<\sum_{j=n+1}^{n+p+1}X_j\}}\right]$$
$$=\sum_{n=1}^{+\infty}E\left[F\left(\sum_{i=1}^{n}\delta_{X_i}, X_{n+1}\right)1_{\{\sum_{i=1}^{n}X_i\leqslant 1<\sum_{i=1}^{n+1}X_i\}}\right]$$

with the convention $\sum_{n+1}^{n}=0$. Since our variables are i.i.d. we can rewrite the left-hand side of our identity to be proved as

$$\sum_{n=1}^{+\infty}\sum_{p=0}^{+\infty}E\left[F\left(\sum_{i=1}^{n-1+p}\delta_{X_i}, X_{n+p}\right)1_{\{S_{n-1}+X_0\leqslant 1<S_n+X_0,S_{n+p-1}\leqslant 1<S_{n+p}\}}\right],$$

or changing variables in n and p,

$$\sum_{n=1}^{+\infty}E\left[F\left(\sum_{i=1}^{n}\delta_{X_i}, X_{n+1}\right)1_{\{S_n\leqslant 1<S_{n+1}\}}\times\left(\sum_{i=0}^{n}1_{\{S_i+X_0\leqslant 1<S_{i+1}+X_0\}}\right)\right].$$

Our identity is proved, the last factor being identically 1 on the subset $\{S_n\leqslant 1<S_{n+1}\}$. $\square$

Since $c(\mu)$ is a normalizing constant it can be easily computed by taking $F\equiv 1$ and we obtain

$$c(\mu)^{-1}=\sum_{n=1}^{+\infty}\frac{1}{n}P(S_n\leqslant 1<S_{n+1}).$$

Taking $F(\eta, y)=\langle 1, \eta\rangle=N$ which is the departure rate (also the number of items in the bin), thus

$$\rho(\mu)=c(\mu)$$

and hence the following result holds.

Corollary 4.2 (formula for the maximal throughput).

$$\rho(\mu)=\frac{1}{\sum_{n=1}^{+\infty}(1/n)P(S_n\leq 1<S_{n+1})}$$

$$=\frac{1}{\sum_{n=2}^{+\infty}(1/n(n-1))P(S_n>1)}=\frac{1}{E(1/N)} \tag{4.1}$$

where $N=\sup\{k\mid S_k\leq 1\}$ *is the time spent by the random walk* $(S_n)_{n\geq 1}$ *below* 1. □

Example. If μ is the uniform distribution on [0, 1], it is well known that the density of S_n restricted to [0, 1] is $x^{n-1}/(n-1)!$ thus $\rho(\mu)=1/(\mathrm{e}-2)$ which is ≈ 1.392. Using Theorem 4.1 it is easy to get in this case the density h of the occupied length of the bin, $h(x)=(\mathrm{e}^x-1)/(\mathrm{e}-2)$.

Remark. If μ_n is the total processing rate when n items are present in the unit, the method can be extended to give the formula

$$\rho(\mu)=\frac{1}{\sum_{n=1}^{+\infty}(1/\mu_n)P(S_n\leq 1<S_{n+1})}.$$

In particular if at most k items can be processed at the same time we obtain,

$$\rho(\mu)=\frac{1}{\sum_{n=1}^{k}(1/n)P(S_n\leq 1<S_{n+1})+(1/k)P(S_{k+1}\leq 1)}.$$

The drawback of the formulas (4.1) is that the estimation of its components related to the random walk $(S_n)_{n\geq 0}$ is almost always impossible. In the next section we will derive some bounds on the throughput using only simple functionals of the distribution μ. We finish with a simple but important proposition which will be used in Section 7, it will permit to compute numerically the throughput in most of the cases.

Proposition 4.3. *If* μ *has no atom and*

$$\hat{\mu}(\gamma)=\int_0^1 \mathrm{e}^{\mathrm{i}\gamma x}\mu(\mathrm{d}x)$$

denotes the Fourier–Laplace transform of μ *then,*

$$\rho(\mu)^{-1}=\mathrm{e}^{\alpha}\lim_{t\to+\infty}\int_0^t 2\,\mathrm{Re}\left(\frac{(1-\hat{\mu}(\gamma+\mathrm{i}\alpha))}{\mathrm{i}(\gamma+\mathrm{i}\alpha)}\log(1-\hat{\mu}(\gamma+\mathrm{i}\alpha))\,\mathrm{e}^{-\mathrm{i}\gamma}\right)\mathrm{d}\gamma$$

with Re(z) *real part of* z *and* $\alpha>0$.

Proof. If we set

$$f(x)=\sum_{n=1}^{+\infty}\frac{1}{n}P(S_n\leq x<S_{n+1}),$$

then f is a continuous function bounded by 1, thus $F(x)=f(x)\,\mathrm{e}^{-\alpha x}1_{\mathbb{R}_+}(x)$ is square integrable. Now if we remark that the Fourier transform of f is $((1-\hat{\mu}(\gamma))/(\mathrm{i}\gamma))\log(1-\hat{\mu}(\gamma))$ and that $f(1)$ is $\rho(\mu)^{-1}$, the Fourier inversion theorem applied to F finishes the proof. $\square$

5. Some bounds on the maximal throughput

An immediate consequence of Corollary 3.2 is the following result.

Proposition 5.1.

$$\rho(\mu)\leqslant 1\Big/\int_0^1 x\mu(\mathrm{d}x)=1/E(X_1). \qquad \square$$

We now proceed to proving lower bounds on the throughput. For this we introduce the measure

$$\mu_\alpha=Z(\alpha)\,\mathrm{e}^{\alpha x}\mu(\mathrm{d}x)$$

where $Z(\alpha)$ is the normalizing factor $1/\int_0^1\mathrm{e}^{\alpha x}\mu(\mathrm{d}x)$.

Proposition 5.2. *The family of $E_{\mu\alpha}(1/N)$ satisfies the inequality*

$$E_{\mu_\alpha}\left(\frac{1}{N}\right)\leqslant E_{\mu_\alpha}\left(\frac{X_1}{\gamma-X_1}\right)+(\gamma-1)\int_0^{+\infty}E_{\mu_\alpha}\left(\frac{1}{N}\right)\mathrm{e}^{-\gamma h}\frac{Z(\alpha+h)}{Z(\alpha)}\,\mathrm{d}h$$

for all $\gamma\geqslant s=\inf\{a\mid P(X>a)=0\}$.

Proof. We start from the formula

$$E_{\mu_\alpha}(\mathrm{e}^{\alpha S_{N+1}}Z(\alpha)^{N+1})=1$$

which follows from the fact that $N+1$ is a regular stopping time for the exponential martingale $(\mathrm{e}^{\alpha S_n}Z(\alpha)^n)_{n\geqslant 1}$.

Multiplying both sides by $Z'(\alpha)Z(\alpha)^{-2}\,\mathrm{e}^{-\alpha\gamma}$ and integrating, we obtain the identity

$$-E\left(\frac{X_1}{\gamma-X_1}\right)=\int_0^{+\infty}\mathrm{e}^{-\alpha\gamma}E(X_1\,\mathrm{e}^{\alpha X_1})\,\mathrm{d}\alpha$$

$$=\int_0^{+\infty}E(\mathrm{e}^{\alpha S_{N+1}}Z(\alpha)^{N-1}Z'(\alpha)\,\mathrm{e}^{-\alpha\gamma})\,\mathrm{d}\alpha$$

$$=-E\left(\frac{1}{N}\right)-\int_0^{+\infty}E\left[\frac{(S_{N+1}-\gamma)}{N}\,\mathrm{e}^{\alpha(S_{N+1}-\gamma)}Z(\alpha)^N\right]\mathrm{d}\alpha.$$

If we notice that $S_{N+1}>1$ a.s., then

$$E\left(\frac{1}{N}\right)\leqslant E\left(\frac{X_1}{\gamma-X_1}\right)+(\gamma-1)\int_0^{+\infty}E_{\mu_\alpha}\left(\frac{1}{N}\right)\mathrm{e}^{-\alpha\gamma}E(\mathrm{e}^{\alpha X_1})\,\mathrm{d}\alpha.$$

Now taking μ_{α_0} instead of μ, we have the same formula and

$$E_{\mu_{\alpha_0}}(e^{\alpha X_1}) = \frac{Z(\alpha+\alpha_0)}{Z(\alpha_0)}. \qquad \square$$

This bound can be used iteratively starting from any a priori bound. An easy a priori estimate is $\rho(\mu) \geqslant k$ if the measure μ is supported by the interval $[0, 1/k]$. In particular we obtain the following corollary.

Corollary 5.3. *The maximum throughput satisfies the inequality*

$$\rho(\mu) \geqslant \sup\left\{ 1 \Big/ E\left(\frac{X_1 + (1/k)(\gamma-1)}{\gamma - X_1}\right) \,\Big|\, \gamma \geqslant 1 \right\}. \qquad \square$$

6. Some properties of the throughput, analysis of examples

6.1. Continuity properties of ρ_μ

The space of distributions on $[0, 1]$ will be endowed with the weak topology, that is

$$\mu_n \to \mu \quad \text{iff} \quad \int f(x)\mu_n(dx) \to \int f(x)\mu(dx)$$

for any continuous function f on $[0, 1]$.

Proposition 6.1. *The function $\mu \to \rho(\mu)$ is continuous at every point μ_0 such that $(\sum_{n=1}^{+\infty} \mu_0^n)\{1\} = 0$, i.e. if the renewal measure associated with μ_0 has no atom at 1 (with μ_0^n denoting the nth power of convolution of μ_0).*

Proof. The assumption $\mu_0^n\{1\} = 0$ ensures that $\mu \to P_\mu(S_n > 1)$ is continuous at μ_0 according to a classical theorem (cf. Billingsley, 1968, Chapter 1). Then the second formula of Corollary 4.2 and Legesgue's theorem are applied to conclude. $\square$

6.2. Behavior when μ converges to δ_0

This is of interest because the size of the items is in general small compared to the size of the storage unit. As in Section 6.1 it is easy to prove that the throughput converges to $+\infty$ when $\mu \to \delta_0$. And if the distributions are shrunk then $\rho(\mu)$ tends to infinity proportionally to $1/E(X_1)$:

If $(X_n)_{n\geqslant 0}$ are i.i.d. random variables on $[0, 1]$ with distribution μ, $x \in]0, 1]$ and ρ_x denotes the throughput associated with the sequence $(xX_n)_{n\geqslant 0}$ then

$$\lim_{x\to 0} xE(X_1)\rho_x = 1.$$

According to Corollary 5.3 and Proposition 5.1,

$$E\left(\frac{xX_1}{1-xX_1}\right)^{-1} \leqslant \rho_x \leqslant \frac{1}{xE(X_1)},$$

using Lebesgue's theorem we get that $xE(X)\rho_x \to 1$ as $x \to 0$.

6.3. The case of the uniform distribution on an interval [a, b]

Using Fourier inversion formula of Proposition 4.3, numerical results are given at the end for various a and b.

Let $(X_n)_{n\geqslant 0}$ a sequence of i.i.d. random variables uniformly distributed on [0, 1], then $(2a(X_n - a + \frac{1}{2}))_{n\geqslant 0}$ are uniformly distributed on $I_a = [\frac{1}{2} - a, \frac{1}{2} + a]$. If ρ_a is the throughput associated with this sequence, then we have the following result.

Proposition 6.2. *The mapping $a \to \rho_a$ is non decreasing on $]0, \frac{1}{2}]$ with*

$$\rho_{1/2} = \frac{1}{\mathrm{e}-2} \quad \text{and} \quad \lim_{a\to 0} \rho_a = \tfrac{4}{3}.$$

Proof. Again using Corollary 4.2 we have,

$$\begin{aligned}\rho_a^{-1} - \sum_{n=2}^{+\infty} \frac{1}{n(n-1)} P\left(\sum_{k=1}^{n} 2a(X_k - a - \tfrac{1}{2}) > 1\right) \\ = \sum_{n=2}^{+\infty} \frac{1}{n(n-1)} P\left(S_n > \frac{1 - n(\frac{1}{2} - a)}{2a}\right).\end{aligned}$$

We deduce that $a \to \rho_a$ is non-decreasing. For $a = \frac{1}{2}$ we have the uniform distribution on [0, 1] which we have already seen and

$$\lim_{a\to 0} \rho_a^{-1} = \tfrac{1}{4} + \sum_{n=3}^{+\infty} \frac{1}{n(n-1)} = \tfrac{3}{4}. \qquad \square$$

In fact $\frac{4}{3}$ is the worst throughput for symmetrical distributions.

Corollary 6.3. *If μ is a symmetrical measure around $\frac{1}{2}$ then $\frac{4}{3} \leqslant \rho(\mu) \leqslant 2$ and the bounds are the best possible.*

Proof. Because of the symmetry around $\frac{1}{2}$, $P(S_2 > 1) \leqslant \frac{1}{2}$, and thus

$$\rho(\mu)^{-1} = \sum_{n=2}^{+\infty} \frac{1}{n(n-1)} P(S_n > 1) \leqslant \tfrac{1}{4} + \sum_{n=3}^{+\infty} \frac{1}{n(n-1)} = \tfrac{3}{4}.$$

Therefore the infimum of the throughput on the symmetrical distributions is $\frac{4}{3}$. Finally 2 is an upper bound according to Proposition 5.1 and it is achieved for $\mu = \delta_{1/2}$. □

Remark. If we consider all the distributions with expectation $\frac{1}{2}$ then 1 is the infimum of the throughput (consider $\mu_n = (1-1/n)\delta_{(1/2+1/(2n))} + (1/n)\delta_{1/(2n)}$ as $n \to +\infty$).

7. Numerical results

(a) Throughput for uniform distribution on $[0, a]0 < a < 1$. See Table 1.

(b) Throughput for the distribution concentrated on $\frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ with parameter r the weight of $\frac{3}{4}$ and mean size a.

$$\rho(a, r) = \frac{-12}{[r^4 + (-16a-2)r^3 + 96a^2r^2 + (-256a^3 + 96a^2 - 4)r + 256a^4 - 256a^3 + 96a^2 - 32a + 2]},$$

$a = 0.5$. See Figure 1.

(c) Mean occupation of the bin for uniform distribution on $[x, x+0.02]$. See Figure 2.

Table 1

a	throughput	bound (5.3)	bound (5.1)
0.05	39.51	38.66	40.00
0.10	19.40	18.65	20.00
0.15	12.69	11.98	13.33
0.20	9.34	8.64	10.00
0.25	7.32	6.63	8.00
0.30	5.98	5.29	6.66
0.35	5.01	4.33	5.71
0.40	4.28	3.60	5.00
0.45	3.71	3.04	4.44
0.50	3.29	2.58	4.00
0.55	2.90	2.21	3.63
0.60	2.54	1.89	3.33
0.65	2.26	1.62	3.07
0.70	2.04	1.38	2.85
0.75	1.87	1.17	2.66
0.80	1.73	1	2.50
0.85	1.62	1	2.35
0.90	1.53	1	2.22
0.95	1.45	1	2.10
1.00	1.39	1	2.00

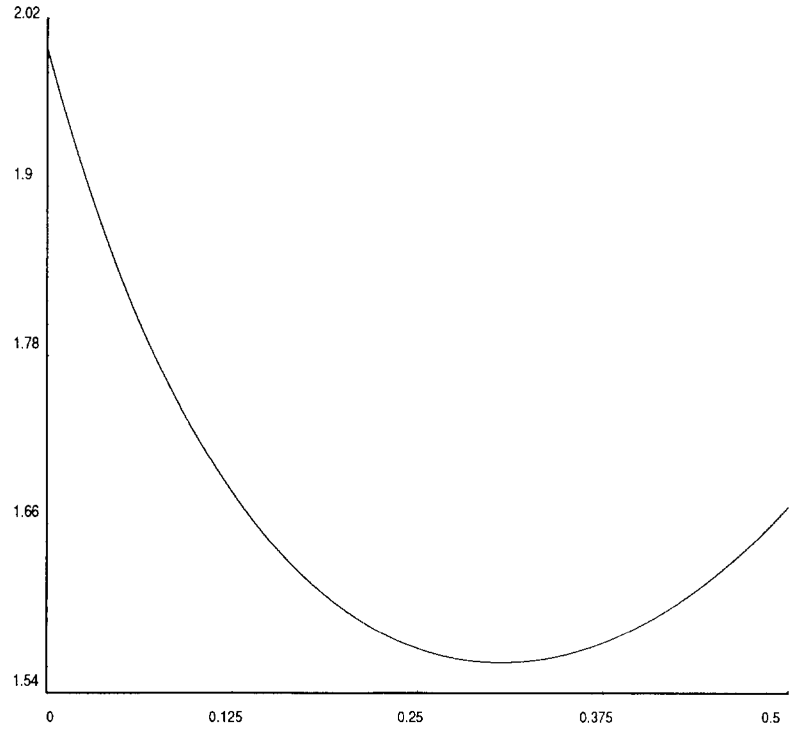

Fig. 1.

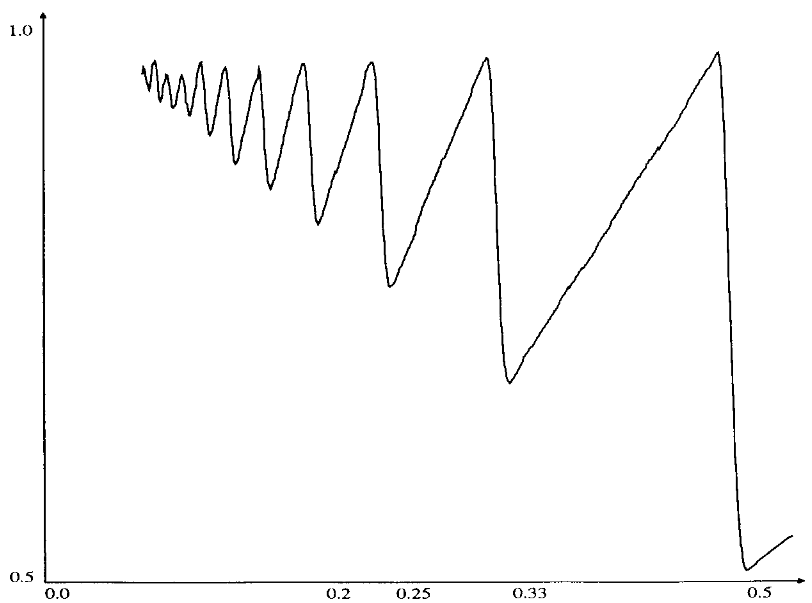

Fig. 2.

References

D. Aldous, Some interesting processes arising as heavy traffic limits in an $M/M/\infty$ storage process, Stochastic Process Appl. 22 (1986) 291-313.

O.I. Aven, E.G. Coffman and Y.A. Kogan, Stochastic analysis of computer storage (Reidel, Dordrecht, 1987).

V.E. Benes, Models and problems of dynamic storage memory allocation, in: R.L. Disney and T.J. Ott, eds., Applied Probability and Computer Science: The Interface, Vol I (Birkhauser, Cambridge, MA, 1982).

P. Billingsley, Convergence of probability measures (Wiley, New York, 1968).

E.G. Coffman, M.R. Garey and D.S. Johnson, Dynamic bin packing, SIAM J. Comput. 12 (1983) 227-258.

E.G. Coffman, T.T. Kadota and L.A. Shepp, A stochastic model of fragmentation in dynamic storage allocation, SIAM J. Comput. 14 (1985) 416-425.

J. Neveu, Courte demonstration du theoreme ergodique sur-additif, Ann. Inst. H. Poincaré 19 (1983).

D. Revuz, Markov Chains (North-Holland, Amsterdam, 1975).

M. Scholl, Analysis of two hash-coded file reorganization policies under steady state conditions, INRIA Report (Rocquencourt Le Chesnay, 1979).